Journal of Information Technology and Digital World (2019)
Vol.01/ No. 02
Pages: 77-86
https://www.irojournals.com/itdw/
DOI: https://doi.org/10.36548/jitdw.2019.2.003

SURVEY ON ACCURACY OF PREDICTIVE BIG DATA ANALYTICS IN HEALTHCARE

Dr. S. Smys,
Professor, Department of Computer Science and Engineering,
RVS Technical Campus,
Coimbatore, India.
Email: smys375@gmail.com

Abstract: The failures in the most of research area, identified that the lack of details about the actionable and the valuable data that conceived actual solutions were the core of the crisis, this was very true in case of the health care industry where even the early diagnoses of a chronic disease could not save a person's life. This because of the impossibility in the prediction of the individual's outcomes in the entire population. The evolving new technologies have changed this scenario leveraging the mobile devices and the internet services such as the sensor network and the smart monitors, enhancing the practical healthcare using the predictive modeling acquiring a deeper individual measures. This affords the researches to go through the huge set of data and identify the patterns along with the trends and delivering solutions improvising the medical care, minimizing the cost and he regulating the health admittance, ensuring the safety of human lives. The paper provides the survey on the predictive big data analysis and accuracy it provides in the health care system.

Keywords: Big Data Analytics, Predictive Modelling, Health Care, Huge Set Of Data, Benefits In Health Care.

1. INTRODUCTION

The health care is main source of economic crisis nowadays as the hospitalization is the most widespread reason for the expenses and does ensure the safety of lives due improper analysis and prediction, so a new health system with the enhanced management, equipped with the recent technologies would offer an rich data set that is necessary in improving the practical heath care system [1].Though the idea behind the big data is not new, it definitions are continuously changing, the different type of definitions significantly characterizing the collection of elements as size, speed along with the type and the complexity allowing one to pursue, implement and discover the new hardware and the software appliance to effectively save , examine and envision the data. The healthcare serves as the fundamental examples, measuring the velocity, variety and the volume of the innate aspect of the data it generated by the field. The data for a heath care is initiated by the different stake holders such as the researchers, health insurers, the government entities etc.

The novel technologies emerged has made possible the capturing of the huge amount of information's associated with the health care (e.g. interactions between the heart rate, respiration and the blood pressure) about each individual patients over the time scale to understand and predict the diseases that require an aggregated approach where the unstructured and the structured data stemming from a high volume of data set are gathered from the clinical and the nonclinical modalities to gain information's about the disease states [2]. So the paper tries to survey the importance of the predictive analytics in the health care system, analyzing the accuracy offered by it in the medical care. The survey was done with according to a systematic search that was carried out in the Google Scholar and the literature relevant to the predictive big data analytics in health were gathered from different scientific database such as the Elsevier, springer, IEEE explorer and other sources in English language was considered and selected for the survey. Totally 25 articles based on the predictive analytics were surveyed, and the primary sources, the strategies followed and the challenges in the adoption and the accuracy in prediction was reviewed. The remaining paper is organized with the predictive analytics in section 2, the uses of big data predictive analytics in section 3, its application and the challenges in adoption and the improvement strategies in section 4 and the conclusion in section 5 before moving into the accuracy achieved in the diagnosis using the predictive analytics let's see what the next section has for us about the basics of the predictive analytics.

2. PREDICTIVE ANALYTICS

The predictive analytics is an advanced form of analytics and has gained repute in the big data related to the health care. It is goes even beyond the data mining, it is a radical branch of data engineering that predicts the existences or the probability according to the availability of the data, it utilizes the data mining method to predict the things that would happen in the future and provides necessary advises from the predictions. The classification and the regression are the two important concepts of the predictive analytics that is composed of variety of statistical and analytical methods. It is capable of handling both the continuous and the discontinuous changes as well. The predictive analysis that is comprised of two words predict and the analysis, so it initially analysis the data provided and predicts the meaning present in it, the flow diagram in the figure.1 below shows the procedure of the prediction. They are the raw data collection, preprocessing, process to transform the preprocessed data into data that is easily handled. This process is usually performed applying the machine learning methods. The learning phase is created using the transformed data and predicts the results using the learning model that was created earlier [3-4].

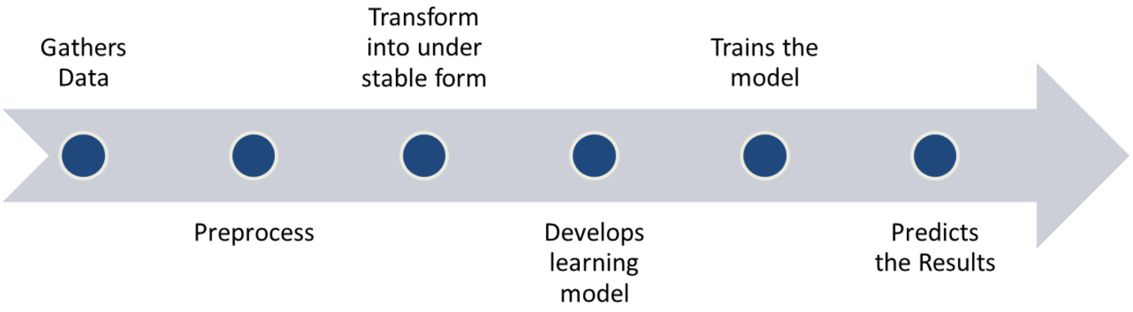


Figure.1 Procedure of Prediction

The models for the predictive analytics are constructed using the data mining tools and the techniques, that identify the hidden patterns or the predictive information's that from the huge volume of data, extracting the valuables available in the data and processing them applying the latest algorithms to detect the hidden information's in them, though there is an palpable acquaintances between the statistics and the data mining the , data mining associated methods have emerged in the respective field other than the statistics [5].

The predictive analytics actually determines the probable future outcome of an event or the likelihood of certain situations that occur in the present, it is totally concerned in predicting the future probabilities and the trends. It automatically analyses the huge set of data involving different variables, some of the methods used in predicting are clustering technique, decision analytics, neural network, decision trees, genetic algorithm, regression modelling, text mining, and hypothesis testing etc. The fundamental element of the predictive analytics is the variable that is coined as the predictor, which is usually measured for a single person or the complete organization in order to predict the future occurrences, the risks and the opportunities hidden within it. The figure.2 below provides the three basic techniques that are used in the predictive analytics [6-10].

DOI: https://doi.org/10.36548/jitdw.2019.2.003

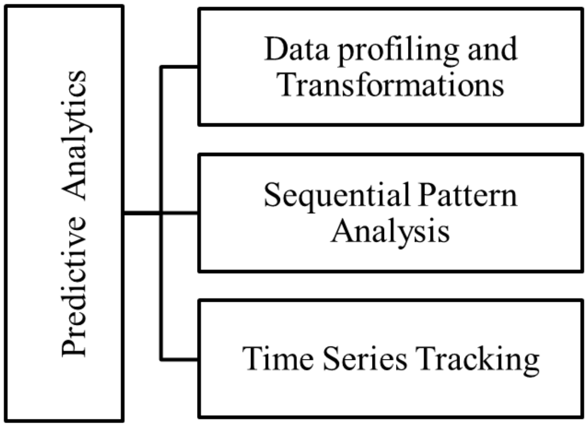
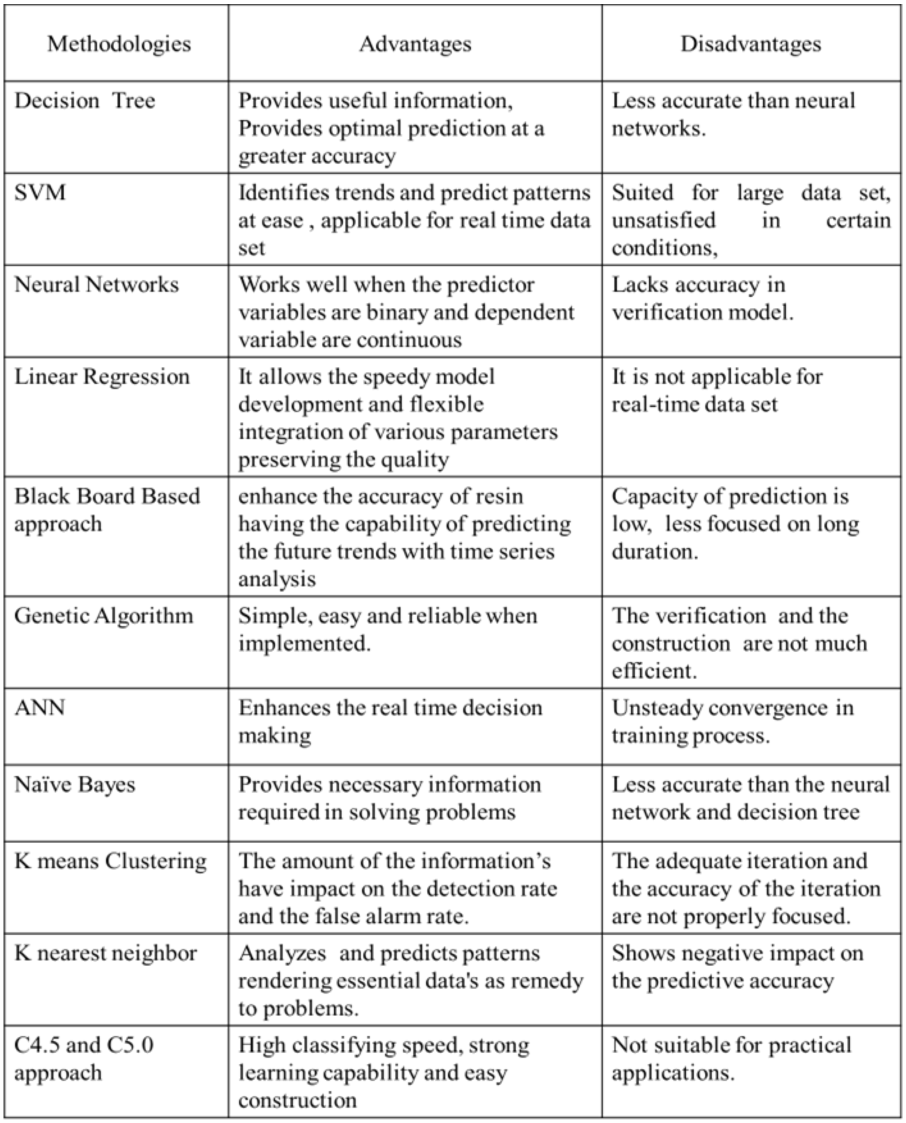

Predictive Analytics
Data profiling and Transformations
Sequential Pattern Analysis
Time Series Tracking

Figure.2 Predictive Analytics techniques

The strategy data profiling and transformation involves the operations and the methods to alter the characteristics of column and the rows, and estimates the dependencies in the aggregated records and the data formats to construct the columns and the rows, the second technique evaluates the relationship in the data in the rows of the data base, it is most probably used in identifying the frequently appearing events across the well-ordered functioning that takes place in a particular instant. The sequence of events arranged with the values defined at different time intervals spaced with the equal distance to offer the conception of the data points that are plotted over time is done by the third type time series analysis. The table.1 below lists out the some of the existing predictive analytics methods and their advantages and disadvantages,



Methodologies	Advantages	Disadvantages
Decision Tree	Provides useful information, Provides optimal prediction at a greater accuracy	Less accurate than neural networks.
SVM	Identifies trends and predict patterns at ease , applicable for real time data set	Suited for large data set, unsatisfied in certain conditions,
Neural Networks	Works well when the predictor variables are binary and dependent variable are continuous	Lacks accuracy in verification model.
Linear Regression	It allows the speedy model development and flexible integration of various parameters preserving the quality	It is not applicable for real-time data set
Black Board Based approach	enhance the accuracy of resin having the capability of predicting the future trends with time series analysis	Capacity of prediction is low, less focused on long duration.
Genetic Algorithm	Simple, easy and reliable when implemented.	The verification and the construction are not much efficient.
ANN	Enhances the real time decision making	Unsteady convergence in training process.
Naïve Bayes	Provides necessary information required in solving problems	Less accurate than the neural network and decision tree
K means Clustering	The amount of the information's have impact on the detection rate and the false alarm rate.	The adequate iteration and the accuracy of the iteration are not properly focused.
K nearest neighbor	Analyzes and predicts patterns rendering essential data's as remedy to problems.	Shows negative impact on the predictive accuracy
C4.5 and C5.0 approach	High classifying speed, strong learning capability and easy construction	Not suitable for practical applications.

Table.1 Methods of Predictive analytics [6] [9] [14] [15] [16] [17] [18] [19]

3. THE USES OF BIG DATA IN PREDICTIVE ANALYTICS IN HEALTH CARE

The big data has a significant role in the predictive analytics especially in the medical domain, the big data are capable of handling a huge set of data that are gathered from the health care sectors, it ensures the solutions to the major issues that arise in the medical industry. The big data analytics in the field of medicine emerges as a promising technology to gather the large quantity of the health care that were accumulated from the patients and the population, and that can be harnessed to progressive prediction, performance, inventions and comparative effectiveness integrating the big data and the future generation analytics into the clinical research, he information reserved are unlimited sources turning out to be knowledgeable information’s, fueling the health care system [20-23]. The figure.3 presents the analytical work flow for the flowing healthcare data.

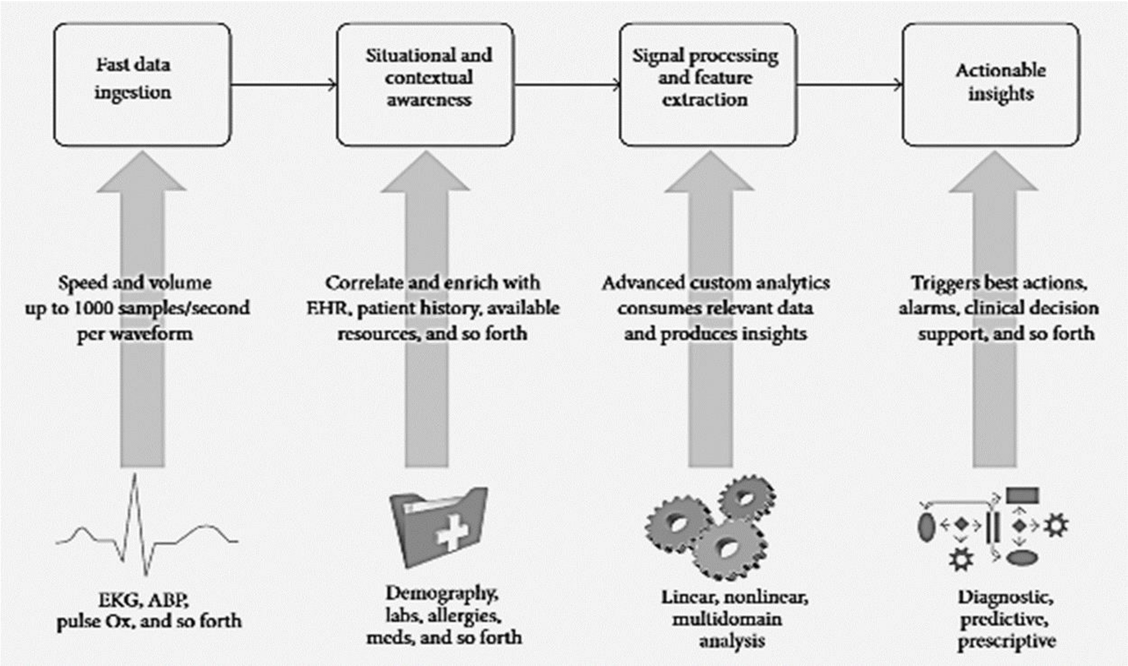
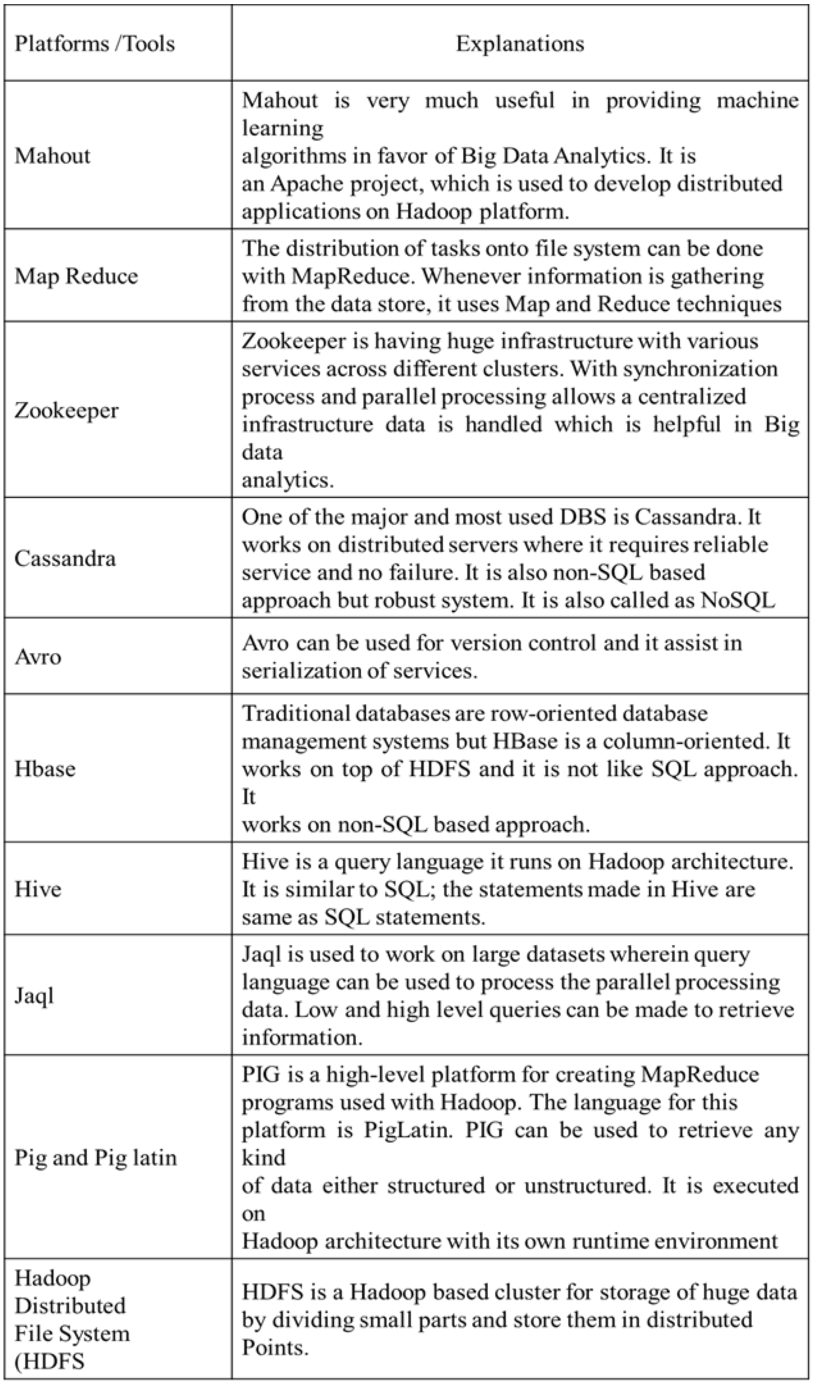

Fast data ingestion
Situational and contextual awareness
Signal processing and feature extraction
Actionable insights
Speed and volume up to 1000 samples/second per waveform
Correlate and enrich with EHR, patient history, available resources, and so forth
Advanced custom analytics consumes relevant data and produces insights
Triggers best actions, alarms, clinical decision support, and so forth
EKG, ABP, pulse Ox, and so forth
Demography, labs, allergies, meds, and so forth
Linear, nonlinear, multidomain analysis
Diagnostic, predictive, prescriptive

Figure. 3 Big Data Analytics for Flowing Health Data [4]

The Table.2below list out some of the platforms and the tools that makes possible the big data predictive analytics in the health care domain. The data from the health care are usually the cluster of genomic data or clinical data or behavioral data, the clinical data is the medical images and the structured and the unstructured information acquired from the medical equipment’s whereas the behavioral data are the data acquired using the mobile app or the sensor engaged in monitoring. The health industry has millions and trillions of patient records that are in heterogeneous form such as the streaming of audio and video, sensor information’s, textual information, written data etc. the information are stored using the database such as the monoDB, Cassandra and H base to provide a security to the information stored and processed in the cloud using the many open source tools, some of the tools and the plat forms used in performing the big data analytics.



Platforms /Tools	Explanations
Mahout	Mahout is very much useful in providing machine learning algorithms in favor of Big Data Analytics. It is an Apache project, which is used to develop distributed applications on Hadoop platform.
Map Reduce	The distribution of tasks onto file system can be done with MapReduce. Whenever information is gathering from the data store, it uses Map and Reduce techniques
Zookeeper	Zookeeper is having huge infrastructure with various services across different clusters. With synchronization process and parallel processing allows a centralized infrastructure data is handled which is helpful in Big data analytics.
Cassandra	One of the major and most used DBS is Cassandra. It works on distributed servers where it requires reliable service and no failure. It is also non-SQL based approach but robust system. It is also called as NoSQL
Avro	Avro can be used for version control and it assist in serialization of services.
Hbase	Traditional databases are row-oriented database management systems but HBase is a column-oriented. It works on top of HDFS and it is not like SQL approach. It works on non-SQL based approach.
Hive	Hive is a query language it runs on Hadoop architecture. It is similar to SQL; the statements made in Hive are same as SQL statements.
Jaql	Jaql is used to work on large datasets wherein query language can be used to process the parallel processing data. Low and high level queries can be made to retrieve information.
Pig and Pig latin	PIG is a high-level platform for creating MapReduce programs used with Hadoop. The language for this platform is PigLatin. PIG can be used to retrieve any kind of data either structured or unstructured. It is executed on Hadoop architecture with its own runtime environment
Hadoop Distributed File System (HDFS	HDFS is a Hadoop based cluster for storage of huge data by dividing small parts and store them in distributed Points.

Table.2 Platforms and Tools of Big Data Analytics [8] [10] [11] [12] [13]

4. CHALLENGES AND ITS APPLICATIONS

The main challenges of the predictive analytics through big data is the data capturing, and storing. They challenges associated with them are the searching the data again and sharing them to the respective persons. It also faces difficulties in the arranging the data after extraction and integrating them, the risk reduction minimizing the errors in the clinical decisions and the other medical aids also needs attention. This could be handled by the development of the digitization of treatment methodologies such as using mobile applications smart sensors etc. the big data in predictive analytics are very much useful in applications of health care, detecting the risk scores of the chronic disease, eluding the patient deterioration, patient self-harm, managing supply chain, ensuring data security, predicting patient mannerisms and prevent the continuous readmission causing the cut down in the revenue losses and speedy access to medical care for the patients[24-25].

5. CONCLUSION

The paper on the survey of the predictive big data analytics in the health care domain, present the general description of the predictive analytics, its techniques and the methods and provides the usefulness of the predictive big data analytics in the health care elaborating the tools that are engaged in making possible the prediction and the also describes the challenges and the applications of it.

In future the paper is to visualize the uses of the and the improvement achieved by the process of prediction in various applications including the health care, and further analyze the accuracy of the different methods of the prediction.

References

[1] Belle, Ashwin, Raghuram Thiagarajan, S. M. Soroushmehr, Fatemeh Navidi, Daniel A. Beard, and Kayvan Najarian. "Big data analytics in healthcare." *BioMed research international* 2015 (2015)..

[2] Schatz, Bruce R. "National surveys of population health: Big data analytics for mobile health monitors." *Big Data* 3, no. 4 (2015): 219-229.

[3] http://www.articlesbase.com/strategic-planning-articles/predictive-analytics-1704860.html

[4] Krumholz, Harlan M. "Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system." *Health Affairs* 33, no. 7 (2014): 1163-1170.

[5] Rumsfeld, John S., Karen E. Joynt, and Thomas M. Maddox. "Big data analytics to improve cardiovascular care: promise and challenges." *Nature Reviews Cardiology* 13, no. 6 (2016): 350.

[6] Jayanthi, N., B. Vijaya Babu, and N. Sambasiva Rao. "Survey on clinical prediction models for diabetes prediction." *Journal of Big Data* 4, no. 1 (2017): 26.

[7] Alharthi, Hana. "Healthcare predictive analytics: An overview with a focus on Saudi Arabia." *Journal of infection and public health* 11, no. 6 (2018): 749-756.

[8] Shyni, S., R. Shantha Mary Joshitta, and L. Arockiam. "Applications of big data analytics for diagnosing diabetic mellitus: issues and challenges." *International Journal of Recent Trends in Engineering & Research* 2, no. 06 (2016): 454-461.

[9] Tekieh, Mohammad Hossein, and Bijan Raahemi. "Importance of data mining in healthcare: a survey." In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 1057-1062. 2015.

[10] Raghupathi, Wullianallur, and Viju Raghupathi. "Big data analytics in healthcare: promise and potential." *Health information science and systems* 2, no. 1 (2014): 3.

[11] Hermon, Rebecca, and Patricia AH Williams. "Big data in healthcare: What is it used for?." (2014).

[12] Dhar, Vasant. "Data science and prediction." *Communications of the ACM* 56, no. 12 (2013): 64-73.

[13] Rajkomar, Alvin, Eyal Oren, Kai Chen, Andrew M. Dai, Nissan Hajaj, Michaela Hardt, Peter J. Liu et al. "Scalable and accurate deep learning with electronic health records." *NPJ Digital Medicine* 1, no. 1 (2018): 18.

[14] Nasridinov, Aziz, Jeong-Yong Byun, Namkyoung Um, and H. Shin. "A study on danger pattern prediction using data mining techniques." *School of Computer Engineering, Dongguk University at Gyeongju, South Korea* (2014).

[15] Bhat, Veena H., Prasanth G. Rao, P. Deepa Shenoy, K. R. Venugopal, and Lalit M. Patnaik. "An efficient prediction model for diabetic database using soft computing techniques." In *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 328-335. Springer, Berlin, Heidelberg, 2009.

[16] Therdphapiyanak, Jakrarin, and Krerk Piromsopa. "An analysis of suitable parameters for efficiently applying K-means clustering to large TCPdump data set using Hadoop framework." In *2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, pp. 1-6. IEEE, 2013.

[17] Kharya, Shweta. "Using data mining techniques for diagnosis and prognosis of cancer disease." *arXiv preprint arXiv:1205.1923* (2012).

[18] Patil, Nilima, Rekha Lathi, and Vidya Chitre. "Comparison of C5. 0 & CART classification algorithms using pruning technique." *Int. J. Eng. Res. Technol* 1, no. 4 (2012): 1-5.

[19] Yue, Jia, Anita Raja, Dingxiang Liu, Xiaoyu Wang, and William Ribarsky. "A Blackboard-based Approach Towards Predictive Analytics." In *AAAI Spring Symposium: Technosocial Predictive Analytics*, vol. 154. 2009.

[20] Smys, S. (2019). BIG DATA BUSINESS ANALYTICS AS A STRATEGIC ASSET FOR HEALTH CARE INDUSTRY. Journal of ISMAC, 1(02), 92-100.

[21] Valanarasu, M. R. (2019). SMART AND SECURE IOT AND AI INTEGRATİON FRAMEWORK FOR HOSPITAL ENVİRONMENT. Journal of ISMAC, 1(03), 172-179.

[22] Raj, Jennifer S. "A COMPREHENSIVE SURVEY ON THE COMPUTATIONAL INTELLIGENCE TECHNIQUES AND ITS APPLICATIONS." Journal of ISMAC 1, no. 03 (2019): 147-159.

[23] Joseph, S. I. T. (2019). SURVEY OF DATA MINING ALGORITHM’S FOR INTELLIGENT COMPUTING SYSTEM. Journal of trends in Computer Science and Smart technology (TCSST), 1(01), 14-24.

[24] Wang, Yichuan, LeeAnn Kung, and Terry Anthony Byrd. "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations." *Technological Forecasting and Social Change* 126 (2018): 3-13.

[25] Bashar, A. (2019). SURVEY ON EVOLVING DEEP LEARNING NEURAL NETWORK ARCHITECTURES. Journal of Artificial Intelligence, 1(02), 73-82.